

Variations d'annotation dans la transcription :

les mises en scène de l'activité sociale dans l'étude du changement linguistique.

Athéna Dupont¹

¹Laboratoire Ligérien de linguistique, Université d'Orléans

Du recueil des données à leur conservation, la chaîne de traitement des corpus oraux comptabilise une soixantaine d'opérations. Parmi celles-ci, la transcription est une étape nécessaire à l'exploitation des données primaires. Dans cette procédure non automatisée entrent en jeu des éléments de perception qui inscrivent socialement les données produites.

Sur le plan méthodologique, Mondada (2002) expose certains des effets structurants de la transcription. Si les conventions sont l'empreinte d'un cadre théorique, leur actualisation construit empiriquement des catégories qui donnent accès à une représentation dynamique de la conversation en cours ; loin de l'illusion de sa neutralité, la transcription incorpore un processus d'évaluation linguistique dans lequel transparaissent les marques de l'agentivité et de la subjectivité. En qualité de locuteur et de sujet social, l'évaluateur projette ses propres représentations sur des données langagières situées.

Le sous-corpus ESLO-Labomedia, constitué dans le cadre d'une thèse en cours sur l'observation des dynamiques sociopragmatiques en jeu dans l'évolution des usages langagiers, procède d'une démarche ethnographique. Les données ont été recueillies dans une association assimilable à un hackerspace par ses activités orientées autour de l'informatique et de l'art numérique. Une présence régulière sur le terrain a permis une contextualisation fine de la situation, qui intègre le profil sociologique des participants. Ainsi, Labomedia réunit des intervenants relativement jeunes et qualifiés, qui disposent d'un solide capital scolaire et culturel.

Comme l'ensemble des modules de la seconde Enquête Sociolinguistique à Orléans (ESLO2), ESLO-Labomedia observe les conventions de transcription en vigueur pour l'ensemble du corpus d'Orléans (Eshkol-Taravella et al., 2011). Ces dernières répondent à une problématique d'accessibilité et de transversalité ; la transcription orthographique permet le

traitement standardisé d'une diversité de situations enregistrées, qui représentent un volume de données conséquent (700h). Chaque document sonore possède trois versions de transcription qui sont construites de manière cumulative. La transcription initiale (transcription, version A) est effectuée à partir du signal, puis modifiée par un second transcripateur (relecture, version B), et éditée par un troisième transcripateur (validation, version C) toujours en lien avec le son. Cette pluralité fait apparaître des différences d'annotation, pour une typologie, voir Hriba dans (Eshkol-Taravella et al., 2011), qui ouvrent la voie à une étude variationniste.

D'autre part, afin de connaître les conditions sociales de production des transcriptions, cette étude est complétée par une série d'entretiens qui a pour visée d'interroger la littéracie des évaluateurs.

Une première étape de l'analyse contrastive consiste à confronter les différentes versions d'une même transcription pour observer les variations de segmentation, notamment dans la structuration des tours de parole et des chevauchements, qui projettent des organisations du discours en interaction, et les variations de perception, qui peuvent être abordées comme des indices d'identification d'usages linguistiques émergents.

Cette étude propose de porter un regard sociolinguistique sur l'indexicalité qui émerge, d'une part, des variations d'annotation repérables dans la transcription, et d'autre part, d'un mode d'expression spécifique enraciné dans son contexte de production, et accessible par l'enquête.

Bibliographie :

ESHKOL-TARAVELLA, I., BAUDE, O., MAUREL, D., HRIBA, L., DUGUA, C., & TELLIER, I. (2011). "Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012", *Traitement Automatique des Langues*, (52-3), 17-46.

GADET, F., & CAPPEAU, P. (2010). "Transcrire, ponctuer, découper l'oral. Bien plus que de simples choix techniques", *Cahiers de linguistique*, (35-1), 187-202.

GOFFMAN, E. (1973). *La mise en scène de la vie quotidienne*, tome 1: *La présentation de soi*. Paris: Minuit.

MAHMOUDIAN, M., MONDADA, L. (1998). (éds) Actes du Colloque "Le travail du chercheur sur le terrain : Questionner les pratiques, les méthodes, les techniques de l'enquête", 13-14.12.1996, *Cahiers de l'ILSL*, (10), Université de Lausanne.

MONDADA, L. (2002). "Pratiques de transcription et effets de catégorisation", *Cahiers de praxématique*, (39), 45-75.

SACKS, H., SCHEGLOFF, E. A., & JEFFERSON, G. (1974). "A simplest systematics for the organization of turn-taking for conversation", *Language*, 696-735.